

Assessing the Accuracy of Satellite-derived Land Cover Classification Using Historical Aerial Photography, Digital Orthophoto Quadrangles, and Airborne Video Data

William G. Kepner¹, Susan M. Skirvin², Stuart E. Marsh², Samuel E. Drake², John K. Maingi³, Curtis M. Edmonds¹, Christopher J. Watts⁴, and David R. Williams⁵

1. U.S. Environmental Protection Agency, Office of Research and Development, National Exposure Research Laboratory, Las Vegas, Nevada
2. Arizona Remote Sensing Center, Office of Arid Lands Studies, University of Arizona, 1955 E. 6th Street, Tucson, Arizona
3. Department of Geography, Morehead State University, Morehead, Kentucky
4. Instituto del Medio Ambiente y el Desarrollo Sustentable del Estado de Sonora (IMADES), Hermosillo, Sonora, Mexico
5. Lockheed Martin Environmental Services, Las Vegas, Nevada

INTRODUCTION

Remotely sensed imagery obtained from Earth-observing satellites now spans decades, making possible the mapping of land cover across large regions by classification of satellite images. However, the accuracy of these derived maps must be known as a condition of the classification. Although the best reference data against which to evaluate classifications are those which have been collected on the ground near the time of satellite overpass, such data are rarely available for retrospective, multi-temporal studies, and alternatives must be found. Based on this need, the U.S. Environmental Protection Agency has established a priority research area for the development and implementation of methods to document the accuracy of classified land cover and land characteristics databases (Jones et al., 2000).

Classification accuracy for four land cover maps of the **San Pedro River Watershed** in southeastern Arizona, USA, and northeastern Sonora, Mexico, were evaluated using aerial photography, digital orthophoto quadrangles, and high resolution airborne video. Classifications were performed at the Instituto del Medio Ambiente y el Desarrollo Sustentable del Estado de Sonora (IMADES, Hermosillo, Mexico) while accuracy was assessed by groups at the University of Arizona (Tucson, Arizona) and Lockheed Martin Environmental Services (Las Vegas, Nevada). This presentation integrates previous work on accuracy assessment (Maingi et al., 1999, Skirvin et al., 2000), describes the methods and results for each classification, and discusses some associated issues.

METHODS

Four sets of land cover maps for the Upper San Pedro Watershed were available for classification accuracy assessment. These maps were generated from 5 June 1973, 10 June 1986, and 2 June 1992 Landsat Multi-Spectral Scanner (North American Landscape Characterization project, USEPA (1993) and 8 June 1997 Landsat Thematic Mapper images). The MSS imagery was projected to Universal Transverse Mercator ground coordinates with a pixel size of 60 meters; the 30 meter TM imagery was resampled and mapped with a pixel size of 60 meters to facilitate comparison.

A 10-class classification scheme was used for the 1973, 1986, 1992 and 1997 digital maps (Table 1). The selected cover classes represent very broad biome-level categories of biological organization and are similar to the ecological formation levels as provided in the Classification System for Biotic Communities of North America (Brown et al., 1979). The ten classes included Forest, Oak Woodland, Mesquite Woodland, Grassland, Desertscrub, Riparian, Agriculture, Urban, Water, and Barren. The classes were selected prior to digitally classifying the imagery and after direct consultation with the major land managers and stakeholder groups within the San Pedro Watershed in Arizona and Mexico (Kepner et al., 2000).

Historical Aerial Photography (1973 and 1986 Maps)

Accuracy assessment of the 1973 and 1986 land cover maps was performed using 1:40,000 scale aerial photography acquired during June 1971, 1983, and 1984. Additionally, April and September frames were utilized from 1972 and 1984, respectively.

Sample points were randomly generated using the registration parameters of digital orthophoto quadrangles acquired in 1992. A grid with a resolution three times the land cover pixel size (180 m by 180 m for the 60-m² pixel) was generated and laid over the land cover map. An allocation of sample points to land cover classes was determined by stratified (by land cover class area) random sampling with a planned sample size of approximately 400 for each year. The samples were apportioned among the land cover classifications according to area with the intention of selecting a minimum of at least 20 samples per class; we assumed a $\pm 5\%$ allowable error and overall map accuracy of at least 75% to calculate the minimum number of sample points at a 95% confidence interval. Class 9 (water) was excluded from the 1986 analysis for lack of data (only 6 possible frames) and does not appear in the final error matrix.

Two mutually exclusive sets of sample points were generated for both 1973 (n = 429) and 1986 (n = 384). Insofar as possible, pixels selected as sample sites represented the center of the 180-m² homogeneous area (i.e., the 3-by-3 square of 9 pixels all representing the same class of land cover) on the land cover map. In cases where a land cover class was not often represented (e.g., water), pixel sample points were chosen with a majority of at least six pixels out of nine in the window belonging to the same class. The land cover classification was based on the center pixel. Aerial photographs from 1971 and 1972 were used to assess the 1973 digital map and from 1983 and 1984 for the 1986 map (Table 2). Once acquired, these photographs were plotted on base maps to establish their positions in relation to the sample points. The digital orthophoto quadrangle imagery was used as a tool to assist in the accurate location of sample points on the photographs during the interpretation process.

When identification of land cover for all sample points was completed, the land cover identification for each sample point was compared with the land cover shown on the digital land cover maps. The classification error matrices and the Producer's, User's and overall classification accuracy along with Cohen's Kappa and Kendall's tau-b coefficients were calculated. To obtain overall map

accuracy, the total of correct classifications (the sum of the major diagonal of the error matrix) is divided by the total number of pixels in the error matrix. The "Producer's" accuracy is calculated by dividing the number of correct classifications in a category by the total number of pixels for that category in the corresponding column (the reference data or photointerpreter class). The "User's" accuracy is calculated by dividing the number of correct classifications in a category by the total number of pixels for that category in the corresponding row (the digital map land cover class).

Digital Orthophoto Quadrangles (1992 Map)

Accuracy assessment of the 1992 land cover map was performed using black and white 1:24,000 scale Digital Orthophoto Quadrangle (DOQ) images acquired from the U.S. Geological Survey. The DOQs were acquired in 1992 and were therefore contemporaneous with the digital land cover map; approximately 60 DOQs were used for the analysis.

We assumed a $\pm 5\%$ allowable error and an overall map accuracy of at least 60% and then used an equation based on binomial probability theory to calculate the minimum number of sample points at a 95% confidence interval.

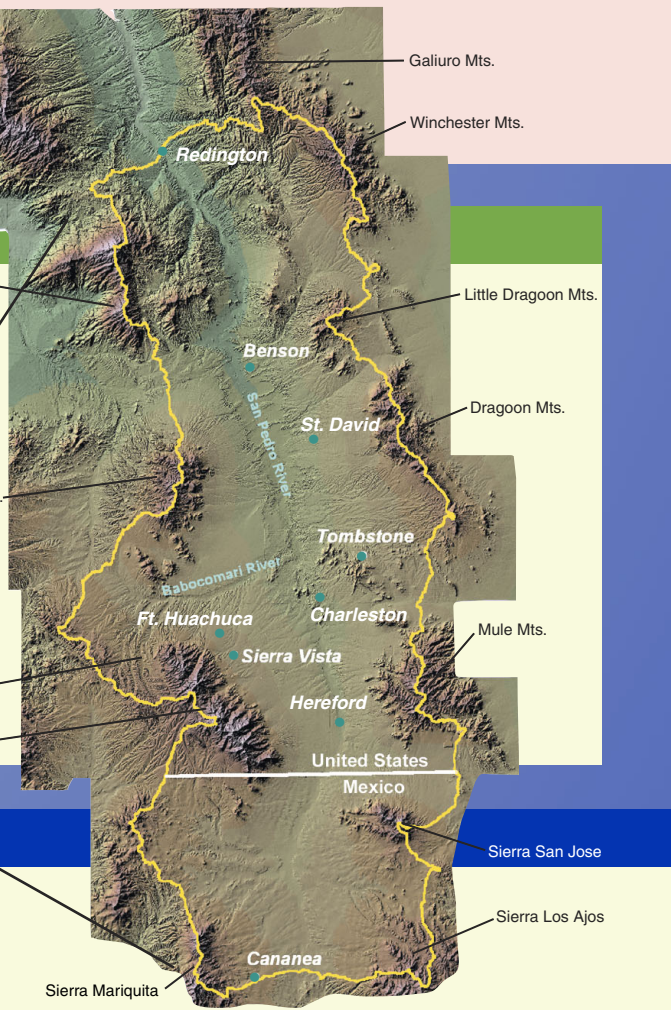
Apportionment of the sample points to the different land cover categories was by stratified random sampling. However, because the area covered by some of the smaller rare land cover classes is negligible compared to the rest of the classes, these classes were not apportioned a sufficient number of sample points. In these situations the smallest sample per class should be 20 or 30 for maps in which the admissible percentage errors are 15% and 10%, respectively (van Genderen and Lock 1977). For this reason, we set 20 as the minimum number of sample points for any class, therefore increasing our total number of required sample points from 370 to 457.

Generation of sample points within the area covered by available DOQs was performed in ERDAS Imagine (Version 8.3) and relied on a window majority rule. In generating each stratified random sample point, a window kernel of 3-by-3 pixels moved across each land cover class and would result in selection of a sample point only if a clear majority threshold of six pixels out of nine in the window belonged to the same class. Generation of sample points in this manner ensured that the points were extracted from areas of relatively homogeneous land cover class.

Most of the land cover classes were fairly well represented in the DOQs. Only the Barren and Agriculture classes were poorly represented in the available DOQs covering 5.3% and 14.2% of the map area, respectively. After interpretation of the 457 DOQ sample frames, the assigned classifications were used to generate an error matrix and used to compute the accuracy of each category, along with both commission and omission errors. In addition, summary statistics for the matrix as a whole (overall classification accuracy, Kappa and tau statistics) were calculated.

Airborne Videography (1997 Map)

Accuracy assessment of the 1997 land cover map was performed using airborne color video data encoded with GPS time and latitude and longitude coordinates. The video data were acquired between 2 and 5 May 1997 and were therefore nearly coincident with the June Landsat TM scene. There were 11 hours of continuously recorded videography of the San Pedro Watershed for the area north of the U.S.-Mexico border. The 1997 video footage was acquired by flying north-south transects spaced 5 km apart and the total flight coverage encompassed a distance of nearly 2000 km. The encoded airborne video GPS time and geographic coordinate data were extracted from the video for each flight line. Flight line coverages were reprojected from geographic to UTM coordinates with the same projection as the 1997 land cover map.



In order to determine which points along the video flight lines fell within homogeneous areas, video point coverages were intersected with the homogeneous land cover polygon coverage. We assumed a $\pm 5\%$ allowable error and an overall map accuracy of at least 60% and determined we needed a total of 369 sample points for the accuracy assessment. Apportionment of sample points to the different land cover classes was stratified according to area. However, because the area covered by some of the smaller land cover classes was negligible compared to the rest of the classes, these classes were not apportioned a sufficient number of sample points. The desired minimum number of sample points for any class was 20, which increased the minimum total number of sample points from 371 to 464.

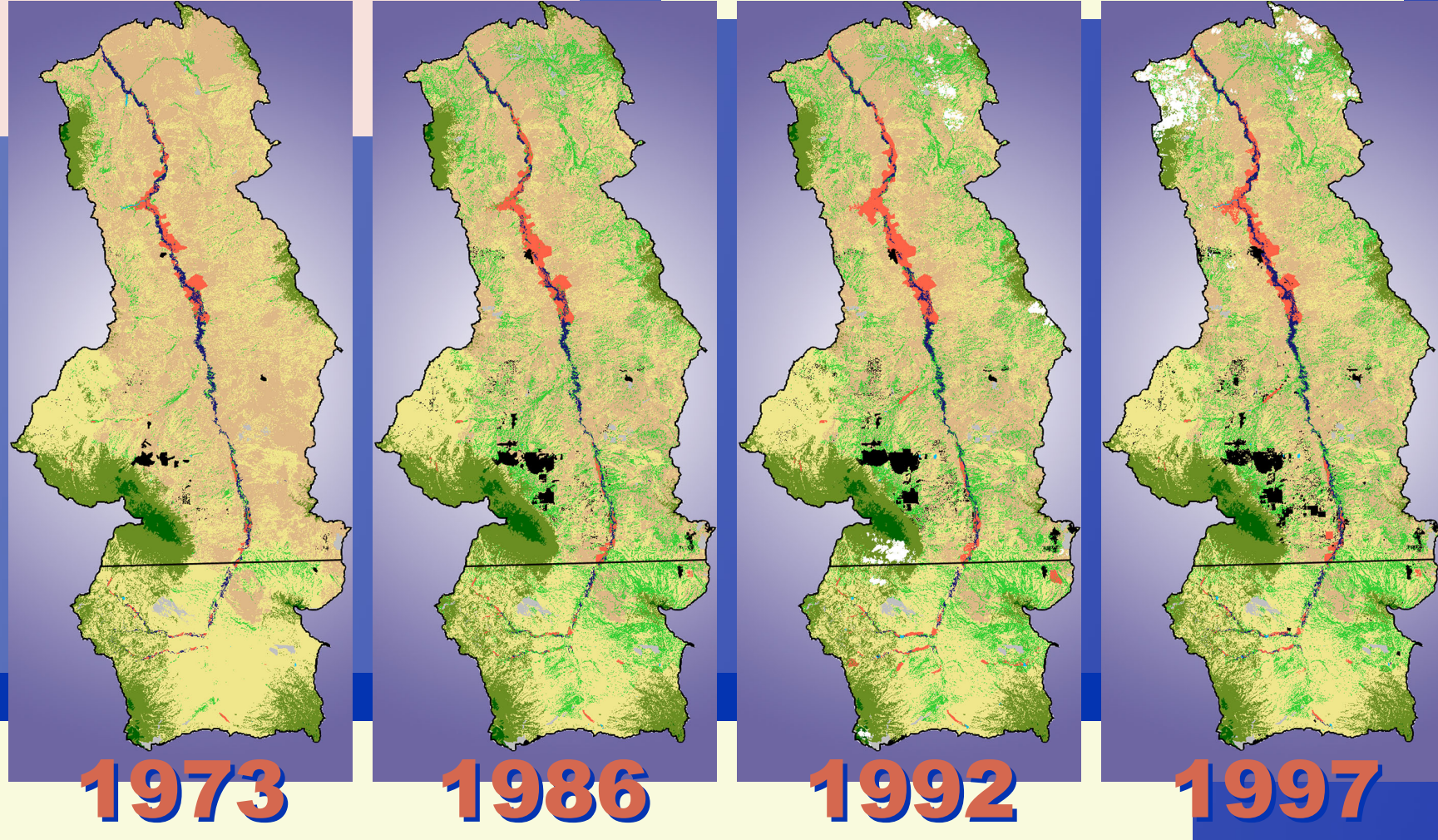
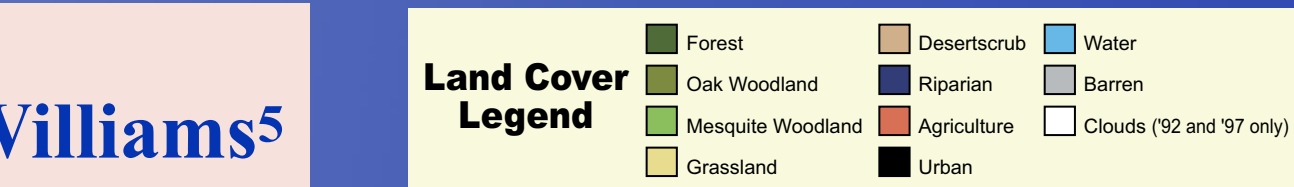
Table 1. Land cover class descriptions for the Upper San Pedro Watershed (U.S./Mexico).

Forest	Vegetative communities comprised principally of coniferous trees potentially over 10 m in height and frequently characterized by closed or multi-layered canopies.
Oak Woodland	Vegetative communities dominated by evergreen trees (<i>Quercus</i> spp.) with a mean height usually between 6 and 15 m. Tree canopy is usually open or interrupted and singularly layered.
Mesquite Woodland	Vegetative communities dominated by leguminous trees with crowns covering more than 15% of the ground and often resulting in dense thickets. Generally found at elevations below 1,200 m.
Grassland	Vegetative communities dominated by perennial and annual grasses with occasional herbaceous species present. Grass height usually under 1 m. Generally found at elevations between 1,100 and 1,700 m.
Desertscrub	Vegetative communities comprised of short shrubs with sparse foliage and small cacti that occur between 700 and 1,500 m in elevation.
Riparian	Vegetative communities adjacent to perennial and intermittent stream reaches. Trees can potentially exceed an overstory height of 10 m and are frequently characterized by closed or multi-layered canopies.
Agriculture	Crops actively cultivated and irrigated (ground and pivot-sprinkler systems).
Urban (Low and High Density)	Land cover dominated by small ejidos (farming villages or communes), retirement homes, or residential neighborhoods (Sierra Vista, AZ). Some heavy industry (open-pit copper mining).
Water	This category is mostly represented by perennial reaches of the San Pedro and Babocomari rivers and includes some attached pools, repressos (earthen reservoirs), tailings ponds, and recreational ponds.
Barren	A cover class represented by large rock outcropping or active and abandoned mines (including tailings) that are largely absent of above-ground vegetation.

Table 2. Film available and acquired for 1973 and 1986 accuracy assessment of digital land cover data for the Upper San Pedro Watershed.

Date	Source	Scale	Film Type	Total Frames
June 21, 1971	USGS/EDC Air Force	1:38,000	Black and White	5
June 29, 1971	USGS/EDC Air Force	1:39,000	Black and White	21
April 21, 1972	USGS	1:40,000	Black and White	41
June 27, 1983	USGS/ NHAP83*	1:40,000	Black and White	11
June 30, 1983	USGS/ NHAP83*	1:40,000	Black and White	2
June 7, 1984	USGS/ NHAP83*	1:40,000	Black and White	22
September 5, 1984	USGS/ NHAP83*	1:40,000	Black and White	6
Total Number of Frames				108

*U.S. Geological Survey National High-Altitude Program



Class 9 (Water) was excluded from analysis for lack of data (only 6 possible frames) and does not appear in the final error matrix. A surplus of about 15% over the calculated minimum number of frames needed in each class was selected, giving the set of 527 frames (instead of the minimum 464).

The latitude, longitude, videotape library identifier, and GPS time for each frame selected were supplied to the videography interpreter, along with the 10 map class descriptions shown in Table 1, and the instruction that each video frame in the selected sample be located, visually interpreted and classified into one of the 10 map classes.

After interpretation and classification of the sampled 527 video frames, the assigned video classifications were used to create an error matrix for the data and to generate Cohen's Kappa and Kendall's tau-b statistics for quantification of overall classification accuracy. Raw percent correct for the matrix, and User's and Producer's accuracies for each class were also calculated.

Table 3. Classification accuracy error matrix for the 1973 digital land cover map using aerial photo interpretations.

	REFERENCE (Aerial Photo Data)										
	1	2	3	4	5	6	7	8	9	10	Grand Total
1973 Digital Land Cover Class	1	19	1	0	0	0	0	0	0	0	20
	2	1	33	0	3	0	0	0	0	0	37
	3	0	1	16	1	0	2	0	0	0	20
	4	0	0	13	92	21	0	0	0	1	128
	5	0	0	0	11	96	0	0	0	0	122
	6	0	0	3	0	2	15	0	0	0	20
	7	0	0	3	0	7	1	10	0	0	22
	8	0	0	0	2	5	0	0	13	0	20
	9	0	0	4	3	6	0	1	0	3	20
	10	0	0	0	2	15	0	0	0	1	20
Grand Total	20	35	53	114	152	18	11	13	5	8	429

Land Cover Class	1973 Map Total	Photo-Interpreter Total	Number Correct	Producer's Accuracy (%)	User's Accuracy (%)
1. Forest	20	20	19	95.00	95.00
2. Oak Woodland	37	35	33	94.29	89.19
3. Mesquite Woodland	20	53	16	30.19	80.00
4. Grassland	128	114	92	80.70	71.88
5. Desertscrub	122	152	96	63.10	78.69
6. Riparian	20	18	15	83.33	75.00
7. Agriculture	22	11	10	90.91	45.45
8. Urban	20	13	13	100.00	65.00
9. Water	20	5	3	60.00	15.00
10. Barren	20	8	2	25.00	10.00
Total	429	429	299		

Overall Accuracy (%) 69.71

Coefficient	Value	Standard Error
Kendall's Tau-B	0.662	0.032
Cohen's Kappa	0.621	0.027

RESULTS

1973/1986 Maps/ Photointerpretation Method

Examination by map class of the 1973 and 1986 results (Tables 3 and 4) indicates that there is substantial variability in classification accuracy. Overall accuracy equaled 69.7% and 68%, respectively for the two digital maps. Producer's and User's accuracies varied considerably, e.g., 1973 Producer's accuracy varied from 100% to 25% and 1986 Producer's accuracy varied from 97% to 0%. Three classes, i.e., Mesquite Woodland, Water, and Barren, presented difficulty for classification. Of these, one is extensive and the other two rare, i.e. <2% overall extent.

1992 Map/DOQ Method

The error matrix showing Producer's and User's, and overall classification accuracy, and including the Kappa and tau coefficients is shown in Table 5. An overall accuracy of about 75% was obtained. Producer's and User's accuracies varied from 100% to 63% and 95% to 44%, respectively. Urban and Barren classes had the lowest User's accuracies, i.e., 44% and 55%.

1997 Map/Airborne Videography Method

Examination by map class of the 1997 results (Table 6) indicates that there is substantial variability in classification accuracy. Producer's accuracy varies from 100% to 54%, and User's accuracy from 100% to 13%. For most classes the two measures are roughly comparable and fall within the range of 60-90%. The Barren, Agriculture, and Mesquite Woodland classes presented some difficulty for classification. User's accuracies reported for these classes equaled 13%, 21%, and 48%, respectively.

Table 4. Classification accuracy error matrix for the 1986 digital land cover map using aerial photo interpretations.

	REFERENCE (Aerial Photo Data)										
	1	2	3	4	5	6	7	8	9	10	Grand Total
1986 Digital Land Cover Class	1	19	1	0	0	0	0	0	0	N/A	20
	2	3	35	0	1	0	0	0	0	N/A	39
	3	0	0	17	3	19	0	1	0	N/A	42
	4	0	0	12	77	12	0	1	0	N/A	104
	5	0	0	8	13	74	0	0	0	N/A	95
	6	0	0	0	1	1	19	2	0	N/A	23
	7	0	0	1	4	3	2	9	1	N/A	20
	8	0	0	0	5	3	0	0	13	N/A	21
	9	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	10	0	0	3	10	7	0	0	0	N/A	20
Grand Total	22	36	41	114	119	21	13	14	N/A	4	384

Land Cover Class	1986 Map Total	Photo-Interpreter Total	Number Correct	Producer's Accuracy (%)	User's Accuracy (%)
1. Forest	20	22	19	86.40	95.00
2. Oak Woodland	39	36	35	97.20	89.70
3. Mesquite Woodland	42	41	17	41.50	40.50
4. Grassland	104	114	77	67.50	74.00
5. Desertscrub	95	119	74	62.20	77.90
6. Riparian	23	21	19	90.50	82.60
7. Agriculture	20	13	9	69.20	45.00
8. Urban	21	14	13	92.90	61.90
9. Water	NA	NA	NA	NA	NA
10. Barren	20	4	0	0.00	0.00
Total	384	384	263		

Overall Accuracy (%) 68.00

Coefficient	Value	Standard Error
Kendall's Tau-B	0.593	0.039
Cohen's Kappa	0.612	0.029

CONCLUSIONS

Overall map accuracies of 68 to 75% were similar among the four maps, despite the use of three different methods and four different reference data sets. For five of the ten land cover classes in all four maps, Producer's and User's accuracy were similar to the overall classification accuracies and ranged between 61 and 100%.

Classes mapped with lower than average accuracy included the small-area Agriculture, Urban, Water and Barren classes and the widespread Mesquite Woodland class. Factors likely to have contributed to class confusions included: cover changes between the dates of image and reference data (e.g., for 1973 and 1986 maps), high spatial variability within classes including areas dominated by soil background, differing interpretations of class definitions by independent assessment teams, and errors in reference data interpretation.

The Agriculture class was mapped with higher Producer's than User's accuracy in all four maps, and was confused with Desertscrub, Riparian and Mesquite Woodland classes. There may have been difficulty in distinguishing fallow and abandoned fields from adjacent Desertscrub and Mesquite Woodland, since their spectral response is dominated by soil background.

The Water class had the smallest areal extent and was likely to have changed between the dates of images and reference data, due to the ephemeral nature of most surface water in this semi-arid environment. Reference aerial photography acquired in June 1983 followed several wetter than usual months resulting from a winter 1982-1983 ENSO event (Easterling et al., 1996; NOAA, 2001). The Water class could not be evaluated in 1986 and 1997 due to insufficient representation in reference aerial photographs and videography.

The Barren class was mapped with poor accuracy overall (0% correct in 1986 map). It was most often confused with Mesquite Woodland, Grassland, and Desertscrub. These classes generally have sparse vegetation cover, with corresponding image pixels dominated by soil or rock

spectral response, and can be difficult to distinguish from truly barren areas at MSS pixel size.

Both User's and Producer's accuracy of all four maps were generally low for Mesquite Woodland (30% and 80% respectively for 1973; 40% to 65% for other years). Class confusions included all but the Forest class, with especially large errors in the Grassland and Desertscrub classes. In large part this may reflect both the spatially and temporally transitional nature of this class, as well as differences in interpretation among the groups performing image classification and assessment. In reality, these classes integrate subtly more often than not, and viable criteria could be established to separate intermediate cases. It is also possible that the resolution of MSS imagery is not adequate to distinguish this class in the heterogeneous watershed environment.

The results discussed in this report indicate that historical aerial photography, DOQ data, and higher resolution data (i.e., airborne video) can be successfully used to perform classification accuracy assessment on land cover maps derived from historical satellite data. The use of DOQ data sets to assess satellite-derived classification accuracies is a viable methodology that can be applied to assess classification accuracy in other project areas using Landsat MSS data obtained from the North American Landscape Characterization program. The use of historic aerial photographs may represent the only available technology for retrospectively assessing pre-1992 digital land cover maps; however, their resolution (1:40,000 scale) often makes this task difficult. The use of georeferenced high-resolution airborne videography appears to provide the best methodology for contemporary landscape characterization projects based on the ease of acquisition and the superior scale resolution (1:200 at 15X-zoom).

Table 6. Classification accuracy error matrix for the 1997 digital land cover map using airborne videography.

	REFERENCE (Video Frame Data)										
	1	2	3	4	5	6	7	8	9	10	Grand Total
1997 Land Cover Class	1	20	4	0	0	0	0	0	N/A	0	24
	2	2	50	0	3	0	0	0	0	N/A	55
	3	0	1	27	13	12	2	0	1	N/A	56
	4	0	8	16	113	21	0	0	1	N/A	159
	5	0	4	4	12	115	0	0	2	N/A	137
	6	0	0	0	0	0	21	2	1	N/A	24
	7	0	0	1	0	15	2	5	1	N/A	24
	8	0	0	0	0	0	0	0	24	N/A	24
	9	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	10	0	0	2	0	19	0	0	0	N/A	24
Grand Total	22	67	50	141	182	25	7	30	N/A	3	527

Land Cover Class	1997 Map Total	Video Total	Number Correct	Producer's Accuracy (%)	User's Accuracy (%)
1. Forest	24	22	20	90.91	83.33
2. Oak Woodland	55	67	50	74.63	90.91
3. Mesquite Woodland	56	50	27	54.00	48.21
4. Grassland	159	141	113	80.14	71.01
5. Desertscrub	137	182	115	63.19	83.94
6. Riparian	24	25	21	84.00	87.50
7. Agriculture	24	7	5	71.43	20.83
8. Urban	24	30	24	80.00	100.00
9. Water	N/A	N/A	N/A	N/A	N/A
10. Barren	24	3	3	100.00	12.50
Total	527	527	378		